

Title	How Machines Force Us to Rethink What It Means to Be Living : Steps to an Existential Robotics
Author(s)	Lestel, Dominique
Citation	Nature Culture. 4 p.38-p.58
Issue Date	2017
oaire:version	VoR
URL	https://doi.org/10.18910/75526
rights	
Note	

Osaka University Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

Osaka University

How Machines Force Us to Rethink What It Means to Be Living¹

Steps to an Existential Robotics

Dominique Lestel

Ecole normale supérieure de Paris

Not to brag, but I'm not very intelligent either.

—Haruki Murakami, *Kafka on the Shore*, 2002/2006.

The post-biological living world into which we are plunging is incredibly complex, for it consists of both organisms and of the ways we manage our relations to them. Today, Westerners use a twofold process to approach the question of the living: first, they try to establish the frontiers between what is living and what is not, and second, they try to establish an objective definition of the living in terms of necessary and sufficient conditions. Proceeding in this fashion seems natural to us, for we have become true Cartesians, but it is nonetheless very problematic. To approach the living in this way means assuming, first, that there exists a class of elements—living beings—that are distinct and

¹ This article was written during a sabbatical year spent at the University of Tokyo with funding from the CNRS, and I wish to thank Sandra Laugier and Philippe Codognot for their help. Translation by Daniela Ginsburg with Emily Sekine, corrected by the author.

spontaneously accessible, whose objective properties we can determine through observation and experimentation. Second, it means assuming that we need only synthesize this class of entities and deduce a few simple principles from it, and we will be able to determine what is living and what is not. But this Cartesian approach (identification, exhaustive definition and recognition of elements that belong to the relevant set) runs into many difficulties: theoretical difficulties, like determining the relevance of the criteria used to define the living; practical difficulties, such as determining whether viruses or crystals are living or not; and logical difficulties, such as knowing if it is legitimate to consider living only what is recognized as such in our own culture. Social robotics—that is, the conception and construction of robots capable of social interaction with humans, animals, or other robots—provides an interesting example in the debate over the frontiers between the living and non-living, transforming itself into a field of inquiry of its own—*existential robotics, which designs 'living' robots*. It presents us with a cognitive conflict between a desire to consider these disturbing artifacts as living agents, and the hegemonic understanding of the living in Western cultures, where biologists have become the authoritative experts on the subject of life.

Thinking the Living from a Second-Person Point of View

Engineers may consider the robots they create to be mere machines—not more than toasters, ultimately—if their own experiences working with these robots suggests otherwise. This position is no more satisfying than the analytic philosophers who want to simply apply the framework of human rights to robots. In the case of certain subjects, reasonable responses turn out to be completely crazy.² My own perspective begins with two philosophical texts that seem to me of particular interest for understanding what is at stake in existential robotics. The first is a note by the American philosopher William James, which lists the reasons why a man could never fall in love with an artificial woman. The second is a seminal article by the British computing pioneer, Alan Turing, in which he presents his famous imitation game—an article so rich that its interpretive possibilities are far from having been exhausted.

In a very concise note published in 1908,³ James establishes a fundamental difference between an actual woman and a hypothetical artificial woman who is

² On this subject, see Günther Anders (2002 [1957]: 34–5): 'My "outrageous" statement is thus no more than a partial reflection of the "exaggeration" that is actually happening today: *it is just an outrageous presentation of what has already been created in exaggeration*'.

³ William James, 'The pragmatist account of truth and its misunderstanders', *Philosophical Review* vol. xvii (January 1908), reprinted in James, *The meaning of truth: a sequel to 'pragmatism'* (2013 [1909]), Chapter VIII.

exactly like a real woman: the latter, he argues, cannot really *care* about another person. She is even less able to care for someone else in a positive way, which prevents him from falling in love with her. At best, the artificial woman can exhibit mere symptoms, which do not correspond to any real consciousness. Thus, James distinguishes artifacts from living beings on the basis of the artifact's inability to care about him—that is, to demonstrate sympathy and recognition, love and admiration. As the reader will see shortly, care for me is not exactly the same as care for the self. And of course, James is not exactly talking about whether or not the artificial woman is *able* to care for me. For him, the operative difference is that the artificial woman does not give me the opportunity to *believe* that she cares about him. What is interesting about James's remark here is that he conceptualizes falling in love from a second-person perspective. The state of being in love is based on my representation of the representations of the other in this case—the artificial woman—has of me. The difficulty in James's argument lies in the fact that 'belief', the concept he mobilizes here, refers to something extremely difficult to objectify.

In this respect, the significance of the 'Turing Test' is that it replaces belief—an inconsistent psychological state that is terribly tricky to grasp—with an interactional **plan**⁴ that can be constantly improved. Turing published the article outlining his famous test of machine intelligence in 1950. Though the Turing Test has been widely discussed over the years, its importance has been underestimated, even by those most passionate about it. Indeed, the test presents a very fertile conceptual staging of intelligence by adopting a co-optative approach (you belong to the club of the intelligent agents if I approve your candidacy)—one very different from the usual logical approaches—and by asking whether a computer, using only symbolic communication, can trick a human into thinking that it, too, is human. The procedure is well-known: a human, H1, poses questions to another human, H2, and to a computer, without knowing which one is which. The three participants are all in different rooms, and H2 and the machine do not communicate with one another. A machine is considered intelligent when H1 takes it for H2 based on the answers it gives. Turing's article emphasizes the necessity of using a purely symbolic relation: 'there [is] little point in trying to make a "thinking machine" more human by dressing it up in [...] artificial flesh' (1950: 434). Turing specifies that the test must take place in conditions of radical abstraction, in which the interrogator is prevented 'from seeing or touching the other competitors, or hearing their voices' (1950: 434). Turing's test deals with a situation of handicap as much as a situation of capacity. We may call this the

⁴ 'Plan' is used here to translate the French word 'dispositif' with the reference to Michel Foucault and Gilles Deleuze.

test's central paradox: an agent's communicational capacity is assessed by placing him/her in a situation of extreme communicational poverty.

Despite these massive handicaps, Turing's test is based on several pivotal changes in how intelligence is viewed:⁵ the shift from defining intelligence to recognizing it; the emphasis on a co-optative approach to intelligence; the recognition of ruse and trickery as central to intelligence; and, above all, the adoption of a second-person conception of intelligence. This second-person perspective is what interests me here, for one of the ambiguities in Turing's article concerns the status attributed to the machine before it is identified as intelligent or not. Namely, what is this 'she/he/it'? The second-person perspective on intelligence solves this difficulty: the machine never says 'I'. The process unfolds appropriately, under the direction of a human being who understands it. One is not *in* dialogue but rather having a monologue *à deux*, until the point that *through* dialogue the machine is revealed to be intelligent. But can the Turing test be generalized to apply to the living *in general*?

James and Turing each approach their subjects in novel and unusual ways. James does so through the possibility of falling in love with an artifact, while Turing does so through the machine's ability to pass for human. Though neither James nor Turing explicitly discuss the question of what life is, together they point the way to a fertile and original approach for thinking life in a radically new way. They provide us with invaluable elements for a concrete, relational, perspectivist, and constructivist approach to the question of what is living. This approach leaves room for some artifacts and does not immediately slam the door shut to the club of the living agents when faced with some slightly exotic entities. This approach is very different from the realist, objectivist, causalist, and essentialist approach taken by the majority of researchers and philosophers who deal with this subject.

A First-Person Point of View on the Living

The two texts by James and Turing I have been discussing lead us to the idea that the question of what is living *for me* is perhaps more relevant than the question of what the living *is* more generally. We may also take it one step further to suggest that what is living for me is who can share *my* existence. The idea of a unique category containing all that can be classified as living, a category that can be objectivized—that is, defined without any reference to a concrete living being—appears highly problematic given the sheer number and diversity of instances in

⁵ On this interpretation of the Turing Test, see Lestel (1996a) and Lestel, Bec & Lemoigne (1993).

which we are willing to recognize the living as living. The problem is the universalist ambition that motivates the search for a general definition of the living that would apply to all its multiple possibilities. Turing and James draw our attention to the fact that when someone identifies an agent as having a somewhat complex property—such as love, intelligence, or life—that person participates fully in the process of identification. Thus, there is a shift from the third-person version of the question—‘What is a living being?’—to the first-person—‘What is a living being *for me*?’ We are led to consider that ultimately what is and is not living is a question that must be asked from the first-person point of view, and not from the third-person perspective, typical of biologists and Artificial Intelligence engineers, who themselves aspire to become kinds of biologists. The advantage of the first-person perspective is that the nature of the agent being evaluated (whether it is natural or artificial) no longer matters. In this view, the set of living beings is the set of all beings recognized as living by at least one other living being.⁶ To adapt Turing and James, a living being is a being I recognize as such (Turing), while a being that I do not recognize as such is not living (James). These two points of view are complementary rather than redundant; from Turing’s perspective, an agent can be considered living as long as I accept that agent as such, even if our so-called experts on life (biologists) are not ready to accept that agent as belonging to the club of the living. And from James’s perspective, an agent that biologists fully accredit as living but who does not convince me, and whom I am not ready to accept as living in my own life, will not be considered living. In ruder words, biologists are only consultants in the matter of the living, and not the ultimate deciders. The space of living beings is not coextensive with the space of biologists’ expertise.

A Second-Person Conception of the Living

If we read more closely, we can see that James’s and Turing’s texts propose a *second-person*, rather than first-person, understanding of the living. In this understanding, I consider you living if I can *experience* you as living in the encounter I have with you, if I can attribute to you a will, a desire, or quite simply a point of view that encounters mine (and this is all the more efficient if your point of view focuses on me). In the second-person perspective, a subject is

⁶ In this view, all prey is living and all that is living is potential prey. The objection that earth is sometimes eaten is not relevant, for it simply means that at least some earth is living. Trickier is the objection that nobody eats an artifact, and thus no artifact is living. But we may reply to this objection that the predator–prey relationship defines an operability of the living but does not subsume all that lives.

constituted through its frictions with interlocutors who consider it, too, to be an interlocutor. Only a few philosophers have discussed this idea; they include, in particular, Martin Buber (1958) (the individual emerges in the I–Thou relationship), Francis Jacques (inter-subjectivity is more important than subjectivity in the dialogical phenomenon, and the presence of absence is essential in that process), and George Herbert Mead (the self only knows itself by mirroring itself in a you, a second-person alter ego). However, all those authors limit themselves to verbal dialogue and communication between humans (or between humans and God).⁷ Opening the second-person perspective to include non-humans in general, and machines in particular, considerably enlarges the scope of this idea, as does shifting from verbal dialogue alone to the encounter between points of view.

The Living as Existential Contract

I have presented constructivist readings of James and Turing. Whereas the American pragmatic philosopher envisioned a somewhat cursory inter-individual encounter, the English mathematician set up an elaborate system to deduce the nature of an artifact. For Turing, the recognition of an intelligent being is concomitant with a somewhat elaborate staging, a point he addresses fully. Recognition takes place through a skillfully orchestrated ‘*mise en demeure*’, a competitive space where the elements that may be taken into account are not limited beforehand. A machine’s intelligence is not a property that a clever philosopher or scientist can discover, but rather an existential contract constructed on the basis of available technologies and the desire to deem the artifact living.

Wanting to have living machines is not, in itself, enough to bring them into being. However, the desire for them to exist is a necessary condition for their creation. Thus, we must think of contemporary existential robotics as lying at the intersection of the conceptualization of what passes for living and what I call libidinal plans which manipulate desire for the living and capture the will to admit the artifact into the club of living beings. The Swiss artist Christian Denisart, whose 2005 piece, *Robots!* plays with robots and human actors, has perfectly grasped this point. He explains that our fascination with robots is tied to the gaze; if we put two wheels and eyes on a coffee machine, he suggests, we will end up becoming attached to it. Denisart holds that the gaze plays an essential role in the ability to give life to an artifact. Moreover, he argues, this life is given through an encounter, implying an exchange of points of view. This attitude is very different

⁷ In the postface to *I and Thou*, Buber (1958) takes a timid step in the direction of the animal, but does not go very far.

from that of classical biology, which bases itself on non-negotiable prerequisites, such as whether or not a being has DNA. The constellation of libidinal cultures, technological virtuosity, and practices of the living allows us to better understand social robots, which both simulate living beings and constitute a trap that captures our desire to ‘make life’ with these beings. Here, ‘simulate’ must be taken to mean, as it does in French, both ‘to act like’ and ‘to be taken for’. The robot’s credibility⁸ thus comes not only from how we conceive of it, but also from its capacity to intoxicate us with our care for it, as well as by its care for us, placing each of us within a relationship of mutual dependence. An artifact becomes living within a complex plan (technical, social, semiotic, psychological) that makes the artifact living *for me*, on the condition that I desire this situation. This desire is not at all a given but is itself already a sophisticated construct. A second-person approach to the living thus connects the capacity to conceive of autonomous agents with a desire to share a life with them.

The Emergence of Libidinal Machines

A key concept for thinking about existential robotics from a constructivist, relational, second-person perspective is that of the *libidinal machine*: a machine that arouses and manipulates desire. To ask whether robots can be considered living in biological terms is the wrong question. The point is not to lead an a priori rebellion against biology, but to give an appropriate response to the question of whether at least some social robots can be considered living. This, I must specify, is not primarily a psychological question. Rather, existential robotics offers us the opportunity to rethink fundamental notions such as life and intelligence, outside the traditional spheres of biology or psychology. According to this point of view, robotics is non-biological and post-psychological. It is non-biological because it puts the field of biology back in its place as the study of carbon-based living beings, and not of the living in general. It is post-psychological because it has no reason to base itself on over-determined cultural categories such as reason or the self⁹ in order to account for the intelligence of machines. A fruitful approach for existential robotics would be to use the difficulties raised as an opportunity to move beyond culturally dated categories, rather than trying to apply them to today’s quasi-autonomous artifacts. In other words, the question is not whether

⁸ I prefer the notion of ‘credibility’ [*vraisemblance*] to that of ‘realism’ because the former is clearly a relational concept, while the latter is an ontological concept that implies we already know what the real is, whereas that is precisely part of the problem.

⁹ There is an entire movement within cognitive psychology that considers robotics very important for studying the functioning of children. To my mind, this is a waste of time and is based on erroneous reasoning.

robots are alive or not, but what devices I can mobilize and what practices I am prepared to engage in to make them so. Existential robotics has always been focused on a somewhat rough harnessing of the desire to grant robots the status of living beings.

The anthropologist Natasha Schull brilliantly explains the concept of libidinal machines, though she herself never uses the term. By analyzing the connivances between the technical characteristics of games of chance in Las Vegas casinos (which are nowadays based on digital technologies) and the phenomenology of players' experiences, Schull (2005) dissects the logic of the libidinal dispositions that artifacts can create in humans. In particular, she breaks down the intimate connection that links states of extreme subjective absorption in the game to procedures used to manipulate space and time so the player will spend as much money as possible on slot machines. The device is created to instill a dissociated subjective state that gamblers call the "zone", in which conventional spatial, bodily, monetary, and temporal parameters are suspended. The zone requires a set of well-identified parameters: being alone, not being interrupted, speed, the ability to choose (even if the procedures for choosing are rigged), and having to follow a certain rhythm (Schull 2005: 73). Speed, especially, is an important aspect of the zone, for that is what, above all, gives players a sense of control. The game is artificially augmented by artifices such as pop-ups and pre-programmed segments, but the player thinks s/he controls the process. Thus, it is easy for players to develop the fantasy that they can get inside the machine (a fantasy perfectly illustrated by the film *Tron*).

In the economy of the zone, money loses its charge as a material means of acquisition and exchange and is converted into the currency of play, a supraeconomic means of suspension from conventional circuits of exchange: 'You're not playing for money; you're playing for credit. Credit so you can sit there longer, which is the goal. It's not about *winning*; it's about *continuing to play*' (Schull 2005: 75).

From the beginning, social robots have generated and manipulated desire; they arouse it and are made to respond to it.¹⁰ To reduce social robots to pure functionalities is to take an impoverished view of them, which also makes it difficult to understand what is really at stake. This is why it is particularly relevant to pay attention to early works of fiction on social robots (Ernest Theodor

¹⁰ This is a point clearly grasped by Emmanuel Grimaud and Zaven Paré (2011). Well before the creation of modern robotics, the American writer Ambrose Bierce (1966) had a similar intuition, placing desire at the center of a curious short story he published in 1910, though the desire he described was that of an artifact and not only of the human who created and interacted with the machine. In Bierce's story, the inventor Moxon is strangled by the machine that has just lost at chess and cannot bear defeat.

Amadeus Hoffmann, Auguste Villiers de L'Isle-Adam, etc.), which explicitly describe machines that arouse desire. As the reader will recall, Frankenstein's problem was precisely that his creature wanted to experience love.¹¹ No determinism imposed that yearning. Other cultures have made different choices—for example, situating intelligent artifacts in the space of mediation between gods, spirits, and demons. In Western culture, art and literature have proven more adept than scientific thought at dealing with the issues raised by social robots. The sociologist Sherry Turkle (2011)—who worries about the fact that we can make a robot say 'I love you'—misses some of the major challenges and goals of robotics by remaining within an overly conventional intellectual space, although she was one of the first to emphasize the identitarian aspect of new information technologies.



The bride of Frankenstein (1935)

Thus, the status of social robots is not as clear-cut as one might imagine. It is a matter not so much of determining whether these creatures are alive, but of recognizing their fundamental ambivalence in the space of the living. Within the

¹¹ In his work *Frankenstein: mythe et philosophie* (1988), the psychoanalyst Jean-Jacques Lecercle shows that Frankenstein's creation of the monster is analogous to a sexual act, and one of the most successful Frankenstein movies is Whale's second, *The bride of Frankenstein* (1935).

perspective I have been discussing, the question is infinitely broader and richer than the question of whether robots can, for example, be held ‘responsible’ or whether we have ethical obligations towards some of them. It is unrealistic to think we can use our conventional, habitual categories from technology and the history of technology to think about these artifacts. Rather, it seems important to me to consider robots as belonging to the category of *meaningful artifacts*, a category generally neglected by historians and philosophers of technology but frequently discussed by anthropologists, ethno-psychiatrists, and art historians. These are artifacts that do not serve to do something but instead contribute to giving meaning and presence to a certain space or situation. These artifacts generate existences¹² or presences rather than procedures, though they can *also* generate procedures. They are artifacts that form particular relations to the living and present a particular view of the living. Treating existential robotics as a matter of engineering replacements for living beings is the wrong path; robots are artifacts that force us to rethink what the living *can* be for us, and this is what makes them so interesting. But the phenomenon of living artifacts did not emerge with robots or computers, as the majority of contemporary theoreticians believe. Living artifacts have a rich and complex history that includes puppets as well as fetishes, charms, and amulets. One of the main characteristics of such artifacts is that they can only function within systems of performance (some of which have been studied in depth by anthropologists). Clearly, this is also characteristic of robots.

The Performances of Existential Robotics

Taking a constructivist, relational position allows us to escape having to choose between an essentialist scientific view and a conventional relativist view, neither of which is satisfying. One of the difficulties in thinking about existential robotics lies in the overly bureaucratic organization of disciplines within universities. Our clear-cut categories of engineering, science, philosophy, anthropology, religion, art, and literature make it difficult to grasp the complexity of what is at stake with robots that/who are personalizing. Existential robotics must be studied at the intersection of multiple disciplinary fields that are typically thought of as separate.

The historian James P. Carse (1987) establishes an interesting opposition between *finite games* and *infinite games*, which is useful to consider here. Finite games are those that end with a winner and a loser, whereas infinite games seek to prolong themselves as long as possible. For him, life itself is the classic example

¹² In this text, I use ‘existence’ in the existentialist rather than ontological sense.

of an infinite game, even if it is one that, paradoxically, the player will always lose. Similarly, a social robot is part of a collaborative game in which humans and machine *play* at giving life to the artifact in a (potentially) infinite way. The decisive issue is how to play with the robot, and in what arenas of performance. The hope is that somehow we will produce something *belonging to* the living (rather than *of* the living), but when the performative aspect of robotics is neglected, we deprive ourselves of the means (that is, the necessary staging) to do that. Existential robotics exists in a state of permanent performance, though the majority of robotics specialists refuse to pay the least bit of attention to this fact.¹³ When robotics specialists do attend to the performative dimension, as does the Japanese roboticist Hiroshi Ishiguro, they overemphasize the spectacle that robots provide and see it as exceptional. But it is not that robots either function normally or as part of a spectacle: robotics *is* spectacle.

Hybrid Communities Negotiating Existence

Within the constructivist and relational approach that I have been describing here, the status of robots is not given from the outset: It evolves gradually on the basis of the existential choices we make and the plan through which we give robots a place. Making some of our artifacts come alive does not only mean transforming them technically so that they conform to a pre-established norm; it also means transforming ourselves so that our existences and the existences of certain artifacts are mutually compatible. In other words, the goal of social robots is to create hybrid communities of shared meaning, interests, and affects¹⁴ with agents whose very status as living is part of what must be negotiated within the community—though this is never the only thing the community must negotiate. It is the question of existence, rather than of life in the strict sense, that is at stake. One way to achieve compatibility is to ‘become machine’ [*s’enmachiner*]. A more fertile, productive approach would be to mobilize engaging fictions that compel us to re-conceptualize the porousness of the space of the living. This effort, in turn, would help us to see the challenges that robots pose as opportunities for rethinking the place of animism in tech-saturated cultures.¹⁵ There are no doubt other approaches we could take as well.

¹³ Cynthia Breazeal’s book *Designing sociable robots* (2002), for example, uses engineering and psychology exclusively, and completely ignores the performative aspect of its star robot Kismet.

¹⁴ On the idea of a hybrid community of shared meaning, interests, and affects, see (Lestel 1996b).

¹⁵ On this subject, see Erik Davis’ intriguing book *TechGnosis: myth, magic, & mysticism in the age of information* (1999). Thierry Bardini has raised this aspect of modern information technology in *Bootstrapping: Douglas Engelbart, coevolution, and the origins of personal computing* (2000) as

To Be Living Is Not a Property but a Plan

Within a constructivist approach to the world, to be living is not a *property* but rather a relational and performative plan that activates skills or capacities without being reducible to them and without having a monopoly on them. Being alive involves a *passive posture* just as much as it mobilizes an *active capacity*. Westerners have great difficulty understanding this basic intuition because they have evolved within a culture that denigrates passivity and promotes action. Puppets and social robots only become living beings if we give them the opportunity to do so, by offering them an ecology in which they can live, and by successfully establishing a form of positive non-intervention with them.¹⁶ This suggestion is necessarily problematic in Western culture, where technology is experienced above all as a space of active participation. But the capacity to make an artifact be alive comes from the capacity to grant it hospitality in *our* living world, to transform ourselves in order to grant the status of living to ‘an other’, and to make our lives mutually compatible. The important idea here is that an artifact’s status as living does not depend solely on what it is or does, but on how we see ourselves and the ways in which we are ready to engage with it. In other words, it is a two-sided effort. It is precisely here that the question of the second-person perspective becomes crucial, for the artifact acquires the status of living being through its ability to engage us with it as if it were living, and to do so in such a way that it becomes living. What makes a Paro¹⁷ alive is not that it tells us it loves us, but that we ourselves come to love it through the affection it lavishes on us and which we return. It becomes living because we feel it to be living, and we feel it to be living because it engages us in such a relation. With it, we construct a space in which the fact that it is alive has meaning.

The crucial point to grasp here is that within a relational and constructivist perspective, there is no difference between affection and the simulation of affection. Indeed, this simulation can become very convincing if we mobilize *ad hoc* performative and relational plan to achieve this. A positivist would object that this is all an illusion. But when a child becomes distraught because someone wants to replace his/her Aibo¹⁸ with a different, identical Aibo, s/he feels negative emotions that are anything but simulated and which affect the child just as any negative emotion would. The key to understanding this lies in a very simple yet underestimated phenomenon: the non-transitivity of the simulation of affects. That

has Fred Turner in *From counterculture to cyberculture: Stewart Brand, the whole earth network, and the rise of digital utopianism* (2006).

¹⁶ The Chinese Taoists had a term for such positive passivity: *wu wei*.

¹⁷ Paro is a very popular companion robot in Japan that has the form of a small seal.

¹⁸ Aibo, a dog-shaped robot built by Sony, has been the most popular companion robot.

is to say, the fact that I can feel very intense affects on the basis of other, purely simulated affects is crucial. Even if the robot has only simulated affects, I can bring it into a technical-psychic trap that generates in me affects that I am not ready to accept as simulated. This leads me to recognize the robot as having a status not quite the same as a dog's, but certainly very different from a toaster's. In other words, I put myself in a position where I recognize the robot as living through its recognition of me. I attribute to it if not life, then at least *a* life, through a second-person perspective realized within a technical, restrictive plan that I have constructed to that end. We may wonder why it is that I need artifacts I consider living—but that is clearly a different question. The relevant point is that, in this way, I obtain a form of living agent that escapes biology. It is not because an artifact is living that we can behave with it the same way we do with living beings; rather, it is because we behave with it as if it were a living being that it ultimately becomes one. The objection of self-hypnosis one might be tempted to raise here thus does not carry much force, because the situation is not one of pure perception, but rather one of lives shared in practice and organized in a concrete world, which is in constant flux.

The Robot's Contaminating Nature

In E.T.A. Hoffmann's story 'The Sandman', one of the least-expected consequences of the automaton Olympia's presence is that humans begin to doubt the nature of other humans surrounding them. Thus the German Romantic writer put his finger on a fundamental point: a robot is a contaminating creature. To be exposed to its presence leads us to doubt the nature of those around us: not *who* they are, but—more disturbingly—*what* they are. And rightly so, if we do not try to hide behind psychological pseudo-explanations. In the end, the robot reveals humanity's little secret: not that humans are pure exteriority, but that their exteriority is constituted exactly in the same way as that of robots, through elaborate social and technical devices. A robot, after all, is a human reduced not only to pure appearance, but, above all, to pure exteriority.

But are humans deeper than artifacts, or are they just more complex surfaces? Instead of getting indignant over such a suggestion, it is more useful to rethink what an appearance is and what a surface is. There is no reason that the latter should have the same meaning when applied to an object as to a living being, for a robot is precisely that intermediate entity for which the notion of surface means both appearance and exteriority; the two meanings emerge fully within a complex and subtle dynamic that the robot is able to arrange between the two. The robot's freedom is very underdetermined, while the human's is clearly

overdetermined. In general, both do what they *must*, and such voluntary subjection to the social order obviously connects them more strongly than any metaphysical characteristics we may identify or manipulate. In other words, La Boétie is more useful for understanding the human/robot convergence than Descartes. Rather than asking if robots have psychological interiority, we should ask if our societies are ready to provide them with a place where they—like the humans who already belong to our societies—can make real choices.

Fictions

Slavoj Žižek (2006), writing on Hegel (in his own inimitable creative way), asks how something like appearance can emerge in the midst of a dumb, flat reality that is *just there*. As Žižek goes on to specify, things do not only appear; they ‘appear to appear’ (2006: 235). And they do so by mobilizing fictions. Fiction is necessary to access the real itself. This is what the Slovenian philosopher perspicaciously observes in the films of Polish director Krzysztof Kieslowski, whose work oscillates between fiction and documentary. Žižek’s conclusion is that at the most radical level, the reality of subjective experience can only be rendered as fiction. His argument is striking, but one of his premises bothers me. Why maintain the opposition between fiction and reality instead of considering, for example, that there are several layers of existence *at once*, more or less real and more or less imaginary? The tension between representations of robots—for example in the press and works of fiction—and what they ‘really’ are is itself part of the ‘robot phenomenon’. Robots are *also* what they seem to be. It is a bit too easy to see the popularizer’s representations as ‘errors’ or ‘abusive simplifications’, a bit too easy to distance oneself from what would from this point of view be obviously uncontrolled ‘drifts’. Robotics is a domain of fantasy, and it is in the very nature of the robot to always be graspable through more or less controlled fictions and more or less accepted imaginaries. This is never the case with, for example, a toaster, which will always be a toaster, no matter what we might dream of one day doing with it (though, in fact, apart from toasting bread, its uses do appear rather limited).¹⁹

Becoming Artifact

Conversely, the figure of the post-human reminds us that it is impossible to understand attempts to render matter living without also understanding the many

¹⁹ Of course, the world is teeming with perverts with unbridled imaginations...

attempts made by humans who long to become matter. For this, we can look to ‘robot fetishists’, humans who seek to behave like robots; Cindy Jackson, the plastic surgery enthusiast, who has made a business out of her ‘dollification’; or Japanese women in Tokyo who invade ‘*ningyō*’ shops, where they have hyper-realistic dolls of themselves made according to the instructions of masters whose reputations as doll-creators are soaring to new heights. The aspiration to become matter is not particular to young women experiencing the identity troubles inherent in adolescence (as well as in Japanese society). Certain scientists discuss similar themes, cloaking them in scientific respectability. The American philosopher Daniel Dennett (1997), for example, argues that we are the ‘direct descendants’ of self-replicating robots with whom living beings originated. Of course, he hastens to add:

it certainly does not follow from the fact that we are descended from robots that we are robots ourselves. After all, we are also direct descendants of fish, and we are not fish; we are direct descendants of bacteria, and we are not bacteria. But [...] we are *made* of robots—or, what comes to the same thing, we are each a collection of trillions of macromolecular machines. And all of these are ultimately descended from the original self-replicating macromolecules. So something made of robots *can* exhibit genuine consciousness, because you do if anything does (Dennett 1997: 24).

Those who *really* wish to work toward becoming robots adopt this conviction, but from a different perspective: if we are all already robots, why not try to become better, higher-performing robots? An early example of such efforts is the cybernetics professor Kevin Warwick of the University of Reading. On Monday, August 24, 1998, at 4 p.m., Warwick shocked the scientific community by implanting an electronic chip (23 mm long and 3 mm in diameter) connected to his nervous system into his arm. He thus became, in his own words, the first Cyborg—that is, a Cybernetic Organism, part human and part machine. Thus, a computer can follow Warwick as he moves about the hallways of the Cybernetics Department; it can open doors, turn on lights, and use computers, all without Warwick lifting a finger. The implant can access all sorts of information, from credit card numbers to blood type and medical records. Such electronic tagging of the human, especially when it involves a sophisticated circuit of silicon chips, constitutes a relatively permanent form of identification. What an individual can do becomes directly linked to his/her status. As Warwick writes, individuals’ work hours, their exact locations within a building, and the company they keep can be determined at all times, so that they can be more easily contacted for messages or urgent meetings. As he adds, such a tool could also be extremely valuable for security purposes. Of course. A second experiment of the same nature took place on March 14, 2002: the project ‘Cyborg 2.0’. A surface of some 100

electrodes was surgically implanted into the median nerve fiber of Warwick's left arm, then connected to a computer by radio signal. The objective was both to connect Warwick's arm to the computer and vice versa. Warwick's wife also participated in the experiment, so they could exchange emotional signals. Of course.

Necessity to Think about Puppets to Understand Robots

Ultimately, robotics is much closer to the traditional art of puppetry than most roboticists and theoreticians are willing to recognize.²⁰ Thinking on marionettes and puppets is more realistic precisely because the artifacts involved have been definitively classified as entirely manipulated objects, and the puppeteer's talent is seen to belong to the exclusive space of artistic representation. However, puppets are, above all, artifacts of the imagination, and for that reason are relevant to those interested in robots. Matthew Cohen (2007) has suggested, that they stimulate spectators and performers to fill in details in their imagination. For George Bernard Shaw (1965: 916), the intensity of marionettes' unchanging facial expressions—an effect human actors cannot achieve—stimulates spectators' imagination. Performers and spectators 'conspire', as Cohen says, in the performance, by attributing the illusion of life to marionettes and 'co-creating' them, so that puppets come to take on the appearance of having their own will. As the historian of puppetry Sally Jane Norman writes:

Puppeteers must struggle and compromise when they animate their recalcitrant figures. They speak of the 'surprises' their puppets keep in store for them: locomotive tics initially seen as handicaps can gradually become essential expressive traits. Kinetic idiosyncrasies are not only functions of material components and structural traits; they can also come from automatisms integrated into the construction of a figure (Norman 1995).

Marionettes are extensions of the human body and operate similarly to the media, according to Marshall MacLuhan; they are a means of communication that an audience amplifies and distorts. They are neither myself nor an other. They become concrete within a disturbing form of object-hood that overflows into personal expression and blends intimate familiarity with foreign otherness.

The possibility of destroying or damaging them reinforces our conviction in their existence. Charles Baudelaire (1995 [1853]) was the first to

²⁰ I presented this idea in a paper I gave in 1993 at the Brussels Free University as part of an international conference on artificial life. I suggested that artificial life should be understood at the intersection of the history of automatons and the history of puppets. This was of course roundly rejected by the engineers in attendance, who see their work as far more serious than the creation of super-puppets.

describe children's confused enthusiasm for destroying their playthings as a way of attempting to see inside to the toy's soul, and their ensuing disappointment as the result of observing the soul's absence. A large part of adult puppetry, from *Ubu* to recent pieces, capitalizes on the marionette's destructive ontology. The marionette invites itself through a performative ontology, but does so clandestinely, unlike robots, which claim life somewhat more aggressively. Gordon Craig's 1917 piece, *Romeo and Juliet: A Motion for Marionettes*, depicts the gradual decomposition over the course of the play of Romeo, who loses his legs, arms, and an eye, and the simultaneous construction of Juliet, who begins as a half-figure with only a stick below the waist but by the end has fully developed into a person, with two legs and two arms. The magic of puppetry operates through a form of mediatized communication in which the object appears to have an independent will that surreptitiously intervenes between the puppeteer-actor's will and the one to whom it is addressed.²¹ Puppets show us that the dichotomy between thing/living being is too simplistic and quite unsatisfying—and not only for children. Sergey Obraztsov dissolved the opposition between inert and living matter by using his own limbs as marionettes, thus giving 'expressive autonomy to the various parts of his body, which were to live their own lives' (Jurkowski 2008: 48). Marionettes, by nature, are silent, and they can be magically destroyed or just as magically reconstructed, in more or less virtuoso performances. The faces of the characters in Wayang theatre are covered with sheets of paper on which protection charms are written to prevent malevolent spirits or tricksters from animating them. The radical opposition between what is alive and what is necessarily inert, between what is living and what is not, is specific to Western culture. Such a strict dichotomy cannot be found elsewhere. From this point of view, robotics is not only an engineering challenge; it is also a veritable cultural provocation, which we must face head-on in order to understand what is at stake.

Conclusion: Expanding the Space of the Categories Mobilized for Conceptualizing the Living

The question of the living does not call for a phenomenological, analytic, or scientific answer that would provide us with the *true* nature of the living, thereby determining once and for all whether or not machines may be part of this exclusive club. I have proposed a constructivist approach that seeks ways of letting machines live and cultivates the desire for them to be alive. It does not make much sense to ask whether or not machines can be alive; the question, rather,

²¹ In his 1597 essay 'Of simulation and dissimulation' (reprinted in 1986 in *The essays*), Sir Francis Bacon provided a model for the kind of ludic dissimulation that characterizes puppetry.

is whether we want them to be alive and if we are ready to give ourselves the means to make that happen. It is necessary to *let machines live*, for they will never be able to unless we create the opportunities for them to do so. The categories one could mobilize to talk about the living must be rethought, in order to provide an operational space that would allow these engaging fictions to constitute *us*, and, at the same time, to establish the nature of those with whom we share our existence. These fictions no doubt destabilize the very notion of existence by causing it to overflow its familiar semantic spaces, but this is a lesser evil. At first glance, some of these fictions appear very strange, but they also achieve a seductive relevance. The fact that the existence of certain agents is transmitted otherwise than through biological life—that is, by contamination rather than by reproduction—gets at the central question raised by existential robotics. We may think of robots as artifacts that trouble the complicities that permeate the interface between life and existence.

Robotics as Performative Engineering for Capturing Affects and Desire

It would be dangerous to directly apply Schull's analysis of libidinal machines to the question of existential robotics, but her work on gambling does lead to a compelling hypothesis. In our reading, we find that at least one form of living can be understood through a technology of dependence on other living beings, while a certain kind of robotics can engineer the capture of affects and desires. Implicit in her text is a spectacular reversal of the logic of the Turing Test, where it is not the one who is co-opted by the observer who may be considered living, but, rather, the one who offers to become the observer's host. Thus, a living being can be thought of as an agent that squats the affectivity of another living being down to its very ontology.

Derrida, in his reading of Mauss, shows the potential of such an approach. In all of anthropology and metaphysics, the gift is always linked to debt—and, more precisely, a debt that cannot be repaid. 'For there to be a gift, *it is necessary* that the donee not give back, amortize, reimburse, acquit himself, enter into a contract, and that he never have contracted a debt' (Derrida 1992: 13). For Derrida, the ideal gift must not appear as a gift. The true gift requires that it be forgotten.

But, after all, what would be a gift that fulfills the condition of the gift, namely, that it not appear as a gift, that it not be, exist, signify, want-to-say as gift? A gift without wanting, without wanting-to-say, an insignificant gift, a gift without intention to give? Why would we still call that a gift? (Derrida 1992: 27)

For Derrida, *only time can be given*. Thus, unlike Mauss, Derrida dissociates the gift from exchange, clearly distinguishing between gifts related to 'having' and gifts related to 'being', by considering how the two forms of gifts mobilize

distinct semantic regimes. Derrida becomes even more interesting for me (without being convincing, however) when he observes that one can only give to a living being. According to Derrida, it makes no sense to say, for example, that one ‘gives’ to the rain, to a mountain, or to a car—which leads him to groundlessly exclude such things from the space of the living. He does not ask the inverse question: *from whom or from what* is it possible to receive? To give existence; to receive existence. The tension there goes to the heart of an ontology of existential contamination beyond species and even beyond biology.

Shattering the Binary between Being Alive/Being Dead

In any case, we lack the vocabulary for approaching the ontology of social robots, which constantly oscillate between the physics of actions, the physiology of elementary components, and the logic of desire. We need, for example, a verb somewhere between ‘to be alive’ and ‘to become living’, which would mean something like ‘to become living in a manner specific to oneself’, and would describe a way of constituting oneself as a particular living being attached to other living beings through the composition of desires rather than through the transition of energies. The Western notion of being alive is overly rigid, and stands in opposition both to being dead (that is, having been alive) and being a thing (never having been alive and never having had the possibility of coming to life). However, being alive does not mean the same thing for a human, a chimpanzee, a slug, a bacterium, a plant, a mushroom, a virus or a robot. We need to invent a verb that indicates the degree of life that has been reached, as well as concepts for thinking the living in a multi-dimensional mosaic space that escapes the overly sharp binary logic we currently rely on. Limit-states of life (fetus, coma, etc.), modalities of life intensity (depression, etc.), and the multiplicity of sites of life (humans, animals, viruses, artifacts, ghosts, etc.) constitute the three areas that explode an overly binary view of the living. In other words, a technology such as existential robotics demonstrates quite clearly that Western thought has become too narrow to conceptualize some of the phenomena emerging within it, and that it must open itself up to other forms of thought: for example, Eastern, Far-Eastern, Australian Aborigine, or Amerindian.

Adopting a Contradictory View of the Living

Thus, we must adopt not only a *pluralist* but also a *contradictory* view of the living: that is, rather than there being a single state of the living that one can decisively identify and define in terms of necessary and sufficient properties—as deeply believed in Western culture—we must accept that there exist multiple phenomena of the living and that some may be based on mutually incompatible

logics. The idea that life can be expressed in very different—even contradictory—modalities is never seriously considered in discussions of robotics. Even Chris Langton, one of the pioneers of artificial life, spoke of life as it could be. He did not imagine generating *lives* so much as *the life* that we privilege, which will always only be one modality among many others. It is thus less the philosophy of existential robotics that interests me, and more the transformations that such strange artifacts as social robots can bring about in our ways of being, of existing, of thinking about the world.

Thinking Beyond Our Species' Capacities

If we want to fruitfully discuss the question of whether artifacts can be alive, we must not only significantly transform our conceptions of 'life' and 'the living', but also our idea of what constitutes a satisfying (or at least partially satisfying) answer to that question. We must become smarter than we are today if we really want to understand what a living being *qua* living being is, without focusing exclusively on a specific set of living beings. We run into two major problems in achieving this, but we are aware of only one of them. We think we can determine what a living being is. But we also think we have exhaustive knowledge of what a living being may be, and we are convinced that we can recognize one when confronted with it. Each of these seemingly simple propositions ultimately poses problems of such complexity that we may never be able to satisfactorily tackle. Scientists have always been convinced that humans represent a form of life—indeed, the only form of life—capable of understanding all past, present, and future forms of life. But this ambition is probably far above the intellectual capacities of the species.

References

- Anders, G. 2002. *L'obsolescence de l'homme*. Paris: Éditions de l'Encyclopédie des nuisances.
- Bacon, F. 1986. *The essays*. (ed. & intro.) J. Pitcher. New York: Penguin Classics.
- Bardini, T. 2000. *Bootstrapping: Douglas Engelbart, coevolution, and the origins of personal computing*. Stanford: Stanford University Press.
- Baudelaire, C. 1995 [1853]. The philosophy of toys. In *The painter of modern life, and other essays* (ed.) J. Mayne, 198–205. London: Phaidon.
- Bierce, A. 1966. Moxon's master. In *The collected works of Ambrose Bierce vol 3: can such things be?*, 88–105. New York: Gordian Press.
- Breazeal, C.L. 2002. *Designing sociable robots*. Cambridge: MIT Press.
- Buber, M. 1958. *I and thou*. (trans.) R.G. Smith. New York: Scribner.

- Carse, J.P. 1987. *Finite and infinite games*. New York: Ballantine Books.
- Cohen, M.I. 2007. Puppetry and the destruction of the object. *Performance Research* **12:4**, 123–31.
- Davis, E. 1997. *TechGnosis: myth, magic & mysticism in the age of information*. New York: Harmony Books.
- Dennett, D. 1997. *Kinds of minds: toward an understanding of consciousness*. New York: BasicBooks.
- Derrida, J. 1992. *Given time, I: counterfeit money*. (trans.) P. Kamuf. Chicago: University of Chicago Press.
- Grimaud, E. and Z. Paré 2011. *Le jour où les robots mangeront des pommes conversations avec un Géminoïd*. Paris: Editions Pétra.
- James, W. 2013 [1909]. *The meaning of truth: a sequel to 'pragmatism'*. New York: Hardpress.
- Jurkowski, H. 2008. *Métamorphose de la marionnette au XXe siècle*. Paris: Editions l'Entretemps.
- Lecercle, J.-J. 1998. *Frankenstein: mythe et philosophie*. Paris: Presses universitaires de France.
- Lestel, D. 1996a. The metaphors of complexity: the language and cognitive resources of artificial life. In *Social Science Information* **35:3**, 511–40. doi:10.1177/053901896035003006.
- 1996b. *L'Animalité. essai sur le statut de l'humain*, Paris: Les Cahiers de L'Herne.
- Lestel, D., L. Bec & J.-L. Lemoigne 1993. Visible characteristics of living systems: aesthetics and artificial life. In *Self-Organization and Life: from Simple Rules to Global Complexity* (Proceeding of the European conference on artificial life), (eds) J.L. Deneubourg et al., .595-603. Brussels: ULB.
- Norman, S.J. 1995. L'art des marionnettes. In *Imagina: actes du forum international des nouvelles images de Monte Carlo*. Monaco: Institut National de l'Audiovisuel/Centre National de la Cinématographie.
- Schull, N.D. 2005. Digital gambling: the coincidence of desire and design. *The Annals of the American Academy of Political and Social Science* **597:1**, 65–81. doi:10.1177/0002716204270435.
- Shaw, G.B. 1965. *The complete prefaces*. London: Hamlyn.
- Turing, A. 1950. Computing machinery and intelligence. *Mind* **59(236)**, 433–60.
- Turkle, S. 2011. *Alone together: why we expect more from technology and less from each other*. New York: Basic Books.
- Turner, F. 2006. *From counterculture to cyberculture: Stewart Brand, the whole earth network, and the rise of digital utopianism*. Chicago: University of Chicago Press.
- Žižek, S. 2006 A plea for a return to difference (with a minor *pro domo sua*) *Critical Inquiry* **32:2**, 226–49.